典型相关分析的理论及其在特征融合中的应用

孙权森[1),2)] 曾生根[1)] 王平安[3)] 夏德深[1)]

1)(南京理工大学计算机系 南京 210094)

2)(济南大学理学院 济南 250022)

3)(香港中文大学计算机科学与工程系 香港)

摘 要 利用典型相关分析的思想,提出了一种基于特征级融合的组合特征抽取新方法.首先,探讨了将典型分析用于模式识别的理论构架,给出了其合理的描述.即先抽取同一模式的两组特征矢量,建立描述两组特征矢量之间相关性的判据准则函数,然后依此准则求取两组典型投影矢量集,通过给定的特征融合策略抽取组合的典型相关特征并用于分类.其次,解决了当两组特征矢量构成的总体协方差矩阵奇异时,典型投影矢量集的求解问题,使之适合于高维小样本的情形,推广了典型相关分析的适用范围.最后,从理论上进一步剖析了该方法之所以能有效地用于识别的内在本质.该方法巧妙地将两组特征矢量之间的相关性特征作为有效判别信息,既达到了信息融合之目的,又消除了特征之间的信息冗余,为两组特征融合用于分类识别提出了新的思路.在肯考迪亚大学CENPARMI手写体阿拉伯数字数据库和FERET人脸图像数据库上的实验结果证实了该方法的有效性和稳定性,而且识别结果优于已有的特征融合方法及基于单一特征进行识别的方法.

关键词 典型相关分析;特征融合;特征抽取;手写体字符识别;人脸识别

中图法分类号 TP391

The Theory of Canonical Correlation Analysis and Its Application to Feature Fusion

SUN Quan-Sen[1),2)] ZENG Sheng-Gen[1)] HENG Pheng-Ann[3)] XIA De-Shen[1)]

1)(*Department of Computer Science, Nanjing University of Science & Technology, Nanjing* 210094)

2)(*School of Science, Jinan University, Jinan* 250022)

3)(*Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong*)

Abstract In this paper, based on feature fusion, a new method of feature extraction is proposed according to the idea of canonical correlation analysis. At first, the framework of canonical correlation analysis(CCA) used in pattern recognition is discussed and its reasonable description is given. This comprises three steps: extracting two sets of feature vectors with the same pattern and establishing the correlation criterion function between the two sets of feature vectors; solving the two sets canonical projective vectors and extracting their canonical correlation features by the CCA algorithm; doing feature fusion for classification by using proposed strategy. Then, the problem of canonical projection vectors is solved when two covariance matrices of training samples are singular. This method is adapted to small sample size and high-dimensional problems, so the applicable range of CCA is extended in theory. At last, the inherent essence of this method used in recognition is analyzed further in theory. The proposed method uses correlation features of two groups of feature vectors as effective discriminant information, so it not only is suitable for information fusion, but also eliminates the redundant information within the features. This is a new

收稿日期:2003-12-19;修改稿收到日期:2005-07-14.本课题得到香港特区政府研究资助局(CUHK/4185/00E)资助.孙权森,男,1963年生,副教授,博士研究生,主要研究方向为模式识别、图像处理、计算机视觉与数据融合.E-mail:qssun@beelink.com.曾生根,男,1977年生,博士,主要研究方向为模式识别、数字图像处理与遥感图像分类.王平安,男,1961年生,博士,教授,研究方向为虚拟现实、计算机图形学与三维医学成像.夏德深,男,1941年生,教授,博士生导师,主要研究方向为图像处理、卫星遥感图像与模式识别.

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

way to classification and recognition. The experiment results on the CENPARMI handwritten Arabic numerals database of Concordia University and FERET face image database show that recognition rate is far higher than that of the algorithm adopting the single feature or the existing fusion algorithm, and that this algorithm is efficient and robust.

Keywords　canonical correlation analysis; feature fusion; feature extraction; handwritten character recognition; face recognition

1 引　言

随着计算机技术的发展，信息融合技术已成为一种新兴的数据处理技术．在信息融合的三个层次(像素级、特征级、决策级)中，以多分类器组合为代表的决策级融合技术已成为当前模式识别领域研究的热点，并在手写体和人脸识别方面取得了较为成功的应用[1~3]，有关特征级融合方面的研究，虽然起步较晚，但已初见端倪[4~6]．

特征级融合的优势是明显的．事实上，对同一模式所抽取的不同特征矢量总是反映模式的不同特性，对它们的优化组合，既保留了参与融合的多特征的有效鉴别信息，又在一定程度上消除了由于主客观因素带来的冗余信息．对分类识别无疑具有重要的意义．

在已有的特征融合方法中，一种是将两组特征首尾相连生成一个新的特征矢量[4]，在更高维的矢量空间进行特征抽取；另一种是利用复向量将同一样本的两组特征矢量合并在一起[5]，在复向量空间进行特征抽取．前一种方法称之为串行融合方法，后一种方法称之为并行融合方法[6]．两种特征融合方法均不同程度地提高了识别率．

典型相关分析(Canonical Correlation Analysis, CCA)是处理两个随机矢量之间相互依赖关系的统计方法，与主成分分析、判别分析一样，在多元统计分析中占有非常重要的地位，是一种很有价值的多元数据处理方法[7,8]．近些年来，国外已开始将CCA用于信号处理、计算机视觉及语音识别等领域[9~12]，并取得了一定进展．而在国内尚未见到此类研究．

本文基于CCA的思想，提出了一种不同于文献[4~6]两种特征融合思路的组合特征抽取方法．首先，探讨了将CCA用于模式识别的理论构架，给出了其合理的描述．即先抽取同一模式的两组特征矢量，建立描述两组特征矢量之间相关性的判据准则函数，然后依此准则求取两组典型投影矢量集，通过给定的特征融合策略抽取组合的典型相关特征并用于分类．其次，解决了当两组特征矢量构成的总体协方差矩阵奇异时，典型投影矢量集的求解问题，使之适合于高维小样本的情形，推广了典型相关分析的适用范围．最后，进一步从理论上剖析了本文方法与文献[4,5]两种特征融合之优劣，指出本文方法之所以能有效地用于识别的内在本质．该方法将两组特征矢量之间的相关性特征作为有效判别信息，既达到了信息融合之目的，又消除了特征之间的信息冗余，为两组特征融合用于分类识别提供了新的途径．

在肯考迪亚大学CENPARMI手写体阿拉伯数字数据库和FERET人脸图像数据库上的实验结果证实了该方法的有效性和稳定性，而且识别率高于基于单一特征的方法进行识别的结果，优于已有的两种特征融合方法．仿真结果也表明了，本文算法不仅可以实现原始特征维数的压缩，而且所抽取的组合的典型相关特征具有良好的分类性能，反映了图像的本质特征．

本文第2节给出将CCA用于组合特征抽取的理论和方法，以及大样本数据库上的实验结果；第3节讨论在高维小样本的情况下，即总体协方差矩阵奇异时，典型投影矢量集的求解问题，给出FERET人脸数据库上的实验结果；第4节从理论上剖析本文方法之所以能有效地用于识别的内在本质；最后给出结论和进一步的研究工作．

2 典型相关分析用于特征融合的理论框架

2.1 CCA的基本思想

在多元统计分析中，经常要研究两组随机变量之间的相关性问题，即把两组随机变量之间的相关性研究化成少数几对变量之间的相关性研究，而且这少数几对变量之间又是不相关的．这种思想是Hotelling在1936年发展起来的[13]．

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

具体来说,给出两个零均值的随机矢量 $\boldsymbol{x}\in R^p$ 与 $\boldsymbol{y}\in R^q$,典型相关分析就是要找到一对投影方向 α 与 β,使得投影 $x_1^*=\alpha^{\mathrm{T}}\boldsymbol{x}$ 与 $y_1^*=\beta^{\mathrm{T}}\boldsymbol{y}$ 之间具有最大的相关性,称这种相关为典型相关,x_1^*,y_1^*为第一对典型变量;继而可以由 $\boldsymbol{x}$,$\boldsymbol{y}$ 出发,找出第二对典型变量 x_2^*,y_2^*,使其与第一对典型变量 x_1^*,y_1^*不相关,且 x_2^*与y_2^*之间又具有最大相关性. 如此下去,使 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的相关性特征提取完毕为止. 这样 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 之间的相关性分析,只需通过分析少数几对典型变量的关系即可达到目的.

一般地,投影方向 α 与 β 可以通过最大化如下的准则函数获得

$$\begin{aligned}\rho &= \frac{E[\alpha^{\mathrm{T}}\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}\beta]}{\sqrt{E[\alpha^{\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\alpha]\cdot E[\beta^{\mathrm{T}}\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}\beta]}}\\ &= \frac{\alpha^{\mathrm{T}}E[\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}]\beta}{\sqrt{\alpha^{\mathrm{T}}E[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}]\alpha\cdot\beta^{\mathrm{T}}E[\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}]\beta}}\\ &= \frac{\alpha^{\mathrm{T}}\boldsymbol{S}_{xy}\beta}{\sqrt{\alpha^{\mathrm{T}}\boldsymbol{S}_{xx}\alpha\cdot\beta^{\mathrm{T}}\boldsymbol{S}_{yy}\beta}}.\end{aligned}$$

这里,$\boldsymbol{S}_{xx}$与 $\boldsymbol{S}_{yy}$ 分别表示 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的协方差矩阵,$\boldsymbol{S}_{xy}$ 表示他们之间的互协方差矩阵.

2.2 组合特征抽取的原理与算法

设 $\omega_1,\omega_2,\cdots,\omega_c$ 为 c 个模式类,模式 ξ 为 n 维实矢量,训练样本空间为 $\Omega=\left\{\xi\mid\xi\in R^n\right\}$. 令 $A=\{\boldsymbol{x}\mid\boldsymbol{x}\in R^p\}$,$B=\{\boldsymbol{y}\mid\boldsymbol{y}\in R^q\}$,$\boldsymbol{x}$ 与 $\boldsymbol{y}$ 分别表示运用不同方法已抽取到的同一模式 ξ 的两组特征矢量. 我们将在变换后的训练样本特征空间 A 与 B 中进行特征融合的讨论.

将 A 与 B 看作是随机矢量空间,$\boldsymbol{x}$ 与 $\boldsymbol{y}$ 是其随机矢量,我们的想法是按照 2.1 节提供的 CCA 的思想,提取 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 之间的典型相关特征,记为 $\alpha_1^{\mathrm{T}}\boldsymbol{x}$ 与 $\beta_1^{\mathrm{T}}\boldsymbol{y}$(第 1 对),$\alpha_2^{\mathrm{T}}\boldsymbol{x}$ 与 $\beta_2^{\mathrm{T}}\boldsymbol{y}$(第 2 对),…,$\alpha_d^{\mathrm{T}}\boldsymbol{x}$ 与 $\beta_d^{\mathrm{T}}\boldsymbol{y}$(第 d 对). 然后分别将 $\alpha_1^{\mathrm{T}}\boldsymbol{x},\alpha_2^{\mathrm{T}}\boldsymbol{x},\cdots,\alpha_d^{\mathrm{T}}\boldsymbol{x}$ 与 $\beta_1^{\mathrm{T}}\boldsymbol{y},\beta_2^{\mathrm{T}}\boldsymbol{y},\cdots,\beta_d^{\mathrm{T}}\boldsymbol{y}$ 看作是变换后的特征分量,即

$$\begin{aligned}\boldsymbol{X}^* &= (\alpha_1^{\mathrm{T}}\boldsymbol{x},\alpha_2^{\mathrm{T}}\boldsymbol{x},\cdots,\alpha_d^{\mathrm{T}}\boldsymbol{x})^{\mathrm{T}}=(\alpha_1,\alpha_2,\cdots,\alpha_d)^{\mathrm{T}}\boldsymbol{x}\\ &= \boldsymbol{W}_x^{\mathrm{T}}\boldsymbol{x};\end{aligned}\tag{1}$$

$$\begin{aligned}\boldsymbol{Y}^* &= (\beta_1^{\mathrm{T}}\boldsymbol{y},\beta_2^{\mathrm{T}}\boldsymbol{y},\cdots,\beta_d^{\mathrm{T}}\boldsymbol{y})^{\mathrm{T}}=(\beta_1,\beta_2,\cdots,\beta_d)^{\mathrm{T}}\boldsymbol{y}\\ &= \boldsymbol{W}_y^{\mathrm{T}}\boldsymbol{y}.\end{aligned}\tag{2}$$

这里 $\boldsymbol{W}_x=(\alpha_1,\alpha_2,\cdots,\alpha_d)$,$\boldsymbol{W}_y=(\beta_1,\beta_2,\cdots,\beta_d)$.

将线性变换

$$\boldsymbol{Z}=\begin{pmatrix}\boldsymbol{W}_x\\ \boldsymbol{W}_y\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}\tag{3}$$

作为投影后的组合特征用于分类. 其中变换矩阵为

$$\boldsymbol{W}=\begin{pmatrix}\boldsymbol{W}_x\\ \boldsymbol{W}_y\end{pmatrix}.$$

定义 1. 称 α_i,β_i 分别为 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的第 i 对典型投影矢量;称 $\alpha_i^{\mathrm{T}}\boldsymbol{x},\beta_i^{\mathrm{T}}\boldsymbol{y}$ 分别为 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的第 i 对典型特征分量. 称 $\boldsymbol{W}$ 为典型投影变换矩阵,$\boldsymbol{Z}$ 为组合的典型相关鉴别特征矢量.

下面具体讨论典型投影矢量与典型相关特征的求法及性质. 为此建立如下准则函数:

$$J(\alpha,\beta)=\frac{\alpha^{\mathrm{T}}\boldsymbol{S}_{xy}\beta}{(\alpha^{\mathrm{T}}\boldsymbol{S}_{xx}\alpha\cdot\beta^{\mathrm{T}}\boldsymbol{S}_{yy}\beta)^{1/2}}\tag{4}$$

这里 $\boldsymbol{S}_{xx}$ 与 $\boldsymbol{S}_{yy}$ 分别表示训练样本空间 A 和 B 的协方差矩阵,$\boldsymbol{S}_{xy}$ 表示它们的互协方差矩阵,且 $\boldsymbol{S}_{xy}^{\mathrm{T}}=\boldsymbol{S}_{yx}$,$r=\mathrm{rank}(\boldsymbol{S}_{xy})$. 本节我们仅考虑 $\boldsymbol{S}_{xx}$ 与 $\boldsymbol{S}_{yy}$ 均非奇异的情况,关于奇异的情况,将在第 3 节专门讨论.

显然,准则函数(4)有如下性质:

(1) $J(k\alpha,l\beta)=J(\alpha,\beta)$. $\forall k,\ l\in$;

(2) $J(\alpha,\beta)$的极值与 α,β 的大小无关,只与 α,β 的方向有关.

为了确保解的唯一性,由上述性质,可令

$$\alpha^{\mathrm{T}}\boldsymbol{S}_{xx}\alpha=\beta^{\mathrm{T}}\boldsymbol{S}_{yy}\beta=1\tag{5}$$

问题变为在约束条件(式(5))下,求使准则函数(式(4))取最大值的典型投影矢量 α 与 β.

将已求出的一对矢量$\left(\alpha_1;\beta_1\right)$作为第 1 对投影矢量;当前 $k-1$ 对投影矢量$\left(\alpha_i;\beta_i\right)$($i=1,2,\cdots,k-1$)取定后,第 k 对投影矢量可通过求解以下最优化问题得到:

$$\text{模型 1}\begin{cases}\max J(\alpha,\beta)\\ \alpha^{\mathrm{T}}\boldsymbol{S}_{xx}\alpha=\beta^{\mathrm{T}}\boldsymbol{S}_{yy}\beta=1\\ \alpha_i^{\mathrm{T}}\boldsymbol{S}_{xx}\alpha=\beta_i^{\mathrm{T}}\boldsymbol{S}_{yy}\beta=0\ (i<k)\\ \alpha\in R^p,\beta\in R^q\end{cases}\tag{6}$$

利用 Lagrange 乘数法,可以将问题转化为求解如下两个广义本征方程的问题:

$$\begin{cases}\boldsymbol{S}_{xy}\boldsymbol{S}_{yy}^{-1}\boldsymbol{S}_{yx}\alpha=\lambda^2\boldsymbol{S}_{xx}\alpha & (7)\\ \boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}\beta=\lambda^2\boldsymbol{S}_{yy}\beta & (8)\end{cases}$$

为了进一步获得满足模型 1 中约束条件的典型投影矢量,令 $\boldsymbol{H}=\boldsymbol{S}_{xx}^{-1/2}\boldsymbol{S}_{xy}\boldsymbol{S}_{yy}^{-1/2}$,应用奇异值分解定理:$\boldsymbol{H}=\sum_{i=1}^{r}\lambda_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}$,其中 $\lambda_1^2\geq\lambda_2^2\geq\cdots\geq\lambda_r^2$是 $G_1=\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}$ 与 $G_2=\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}$的非零本征值,$\boldsymbol{u}_i$与 $\boldsymbol{v}_i$ 分别为 G_1 与 G_2 对应于 λ_i^2的单位正交的本征矢量. 这里,$i=1,2,\cdots,r$.

关于模型 1 的最优解及有关的性质,我们可以归纳为如下的定理(详细的推导过程可参阅文献[7,8]).

定理 1. 令 $\alpha_i=\boldsymbol{S}_{xx}^{-1/2}\boldsymbol{u}_i$,$\beta_i=\boldsymbol{S}_{yy}^{-1/2}\boldsymbol{v}_i$,$i=1,2,\cdots,r$. 则有

(1) $\boldsymbol{G}_1$ 与 $\boldsymbol{S}_{yy}^{-1}\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}$($\boldsymbol{G}_2$ 与 $\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}\boldsymbol{S}_{yy}^{-1}\boldsymbol{S}_{yx}$)有

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

相同的非零本征值，且满足 $1 \geq \lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_r^2 > 0$ $(r = \mathrm{rank}(\boldsymbol{S}_{xy}))$；

(2) α_i 与 β_i 分别是广义本征方程(7)与(8)对应于本征值 λ_i^2 的本征矢量．且 $\alpha_i = \lambda_i^{-1}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}\beta_i$，$\beta_i = \lambda_i^{-1}\boldsymbol{S}_{yy}^{-1}\boldsymbol{S}_{yx}\alpha_i$；

(3)
$$\begin{cases} \alpha_i^{\mathrm{T}}\boldsymbol{S}_{xx}\alpha_j = \beta_i^{\mathrm{T}}\boldsymbol{S}_{yy}\beta_j = \delta_{ij} \\ \alpha_i^{\mathrm{T}}\boldsymbol{S}_{xy}\beta_j = \lambda_i\delta_{ij} \end{cases} \tag{9}$$

其中 $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$，$i, j = 1, 2, \cdots, r$.

定理 1 实际上给出了满足约束条件(式(5)，(6))且使目标函数(式(4))达到极值的解，也即模型 1 的最优解.

推论 1.　若上述矩阵 $\boldsymbol{G}_1$ 或 $\boldsymbol{G}_2$ 的全部特征值为 $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_r^2 > \lambda_{r+1}^2 = \cdots \lambda_p^2 = 0(p \leq q)$，则鉴别准则函数

$$J(\alpha_i, \beta_i) = \lambda_i (i = 1, 2, \cdots, p).$$

可见，当矩阵 $\boldsymbol{G}_1$ 或 $\boldsymbol{G}_2$ 的本征值为 $\lambda_j^2 (j = r+1, \cdots, p)$ 时，$J(\alpha_i, \beta_i) = 0$. 此时不能抽取到有效的典型相关特征. 由此不难推出如下结论.

定理 2.　在准则函数(4)下，满足约束条件(5)和(6)的有效的典型投影矢量的个数最多为 r 对($r = \mathrm{rank}(\boldsymbol{S}_{xy})$)，且所取的 $d(\leq r)$ 对典型投影矢量可分别由式(7)与(8)两个广义特征方程的前 d 个最大本征值所对应的满足(9)的本征矢量构成.

定理 3.　在上述特征融合策略(3)下，所抽取的组合特征是不相关的，且组合的投影矢量关于 $\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}$ 的协方差矩阵 $\boldsymbol{S}$ 共轭正交．这里

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{xx} & \boldsymbol{S}_{xy} \\ \boldsymbol{S}_{yx} & \boldsymbol{S}_{yy} \end{pmatrix}.$$

证明．令 $\boldsymbol{X}^* = \boldsymbol{W}_x^{\mathrm{T}}\boldsymbol{x} = (x_1^*, x_2^*, \cdots, x_d^*)^{\mathrm{T}}$，$\boldsymbol{Y}^* = \boldsymbol{W}_y^{\mathrm{T}}\boldsymbol{x} = (y_1^*, y_2^*, \cdots, y_d^*)^{\mathrm{T}}$. 从定理 1 知，对 $\forall i \neq j$，有

$$\mathrm{cov}(x_i^*, x_j^*) = \alpha_i^{\mathrm{T}}\boldsymbol{S}_{xx}\alpha_j = 0,$$
$$\mathrm{cov}(y_i^*, y_j^*) = \beta_i^{\mathrm{T}}\boldsymbol{S}_{yy}\beta_j = 0,$$
$$\mathrm{cov}(x_i^*, y_j^*) = \alpha_i^{\mathrm{T}}\boldsymbol{S}_{xy}\beta_j = 0.$$

因此

$$\mathrm{cov}(x_i^* + y_i^*, x_j^* + y_j^*) = \mathrm{cov}(x_i^*, x_j^*) + 2\mathrm{cov}(x_i^*, y_j^*) + \mathrm{cov}(y_i^*, y_j^*) = 0.$$

也就是所抽取的组合特征 Z 的分量之间是不相关的. 又因为

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \boldsymbol{S}_{xx} & \boldsymbol{S}_{xy} \\ \boldsymbol{S}_{yx} & \boldsymbol{S}_{yy} \end{pmatrix} \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} = \alpha_i^{\mathrm{T}}\boldsymbol{S}_{xx}\alpha_j + \alpha_i^{\mathrm{T}}\boldsymbol{S}_{xy}\beta_j + \beta_i^{\mathrm{T}}\boldsymbol{S}_{yx}\alpha_j + \beta_i^{\mathrm{T}}\boldsymbol{S}_{yy}\beta_j = 2(1 + \lambda_i)\delta_{ij}.$$

所以 $\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix}$ 与 $\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix}$ 是关于 S 共轭正交的，这里 $i, j = 1, 2, \cdots, r$.　证毕.

2.3　算法步骤

1. 先抽取同一模式的两组不同特征矢量，构成原模式变换后的训练样本空间 A 与 B；

2. 计算 A 与 B 中样本的总体协方差矩阵 $\boldsymbol{S}_{xx}$，$\boldsymbol{S}_{yy}$ 及互协方差矩阵 $\boldsymbol{S}_{xy}$；

3. 根据定理 1，计算 $\boldsymbol{G}_1$ 与 $\boldsymbol{G}_2$ 的非零本征值 $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_r^2$ 以及对应的标准正交的本征矢量 $\boldsymbol{u}_i$，$\boldsymbol{v}_i (i = 1, 2, \cdots, r)$；

4. 根据定理 1 计算所有的典型投影矢量 α_i 与 $\beta_i (i = 1, 2, \cdots, r)$. 取前 d 对投影矢量构成变换矩阵 $\boldsymbol{W}_x$ 与 $\boldsymbol{W}_y$；

5. 利用线性变换(式(3))抽取组合特征并用于分类.

由定理 1 中的(2)可知，在实际求解过程中，只需求出每对典型投影矢量中的一个，另一个可以通过它们之间的关系获得. 一般情况下，可选择阶数低的($\min(p, q)$，不妨设 $p \leq q$)矩阵 $\boldsymbol{G}_1$ 来计算其本征值和本征矢量，这样可减少计算量.

2.4　实验 1

本实验旨在检验本文算法在大样本情况下的有效性. 实验采用国际上广泛使用的肯考迪亚大学 CENPARMI 手写体阿拉伯数字数据库，该数据库包含 0~9 共 10 类手写体阿拉伯数字样本，其中每类的训练样本为 400 个，测试样本为 200 个，这样总共有 4000 个训练样本和 2000 个测试样本. 文献[14]通过对手写体数字图像作预处理，已提取出如下 4 个图像特征：

X^{G}——256 维的 Gabor 变换特征；

X^{L}——121 维的 Legendre 矩特征；

X^{P}——36 维的 Pseudo-Zernike 矩特征；

X^{Z}——30 维的 Zernike 矩特征.

将上述 4 种特征两两组合，在训练样本空间分别按照本文算法(2.3 节)求取相应的典型投影矢量集，并通过线性变换(3)抽取组合的典型相关鉴别特征. 采用最小距离分类器，分类结果见表 1.

这里在求取典型投影矢量集时，应先求维数低的投影矢量，再根据定理 1 中的(2)求取维数高的投影矢量. 如组合特征 X^{G} 与 X^{P} 时，只需在 36 阶的方阵上先计算其本征值和相应的本征矢量，就可以获得 36 维和 256 维的典型投影矢量.

为了与已有的两种特征融合方法进行比较，这里也给出了相应的实验结果. 首先将每类样本矢量归一化，然后按照文献[4]的融合方法，将同一样本的两组特征矢量合并为一个高维矢量；再按照文献[5,6]提供的融合方法，利用复向量将同一样本的两

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

组特征组合到一起；最后分别在组合后的空间上以训练样本的总体散布矩阵为产生矩阵，利用 K-L 变换，进行组合特征抽取. 在最小距离分类器下，其分类错误率见表 1.

从表 1 可以看出，采用本文提出的方法无论组合哪两类特征，识别错误率均低于对应的串行融合与并行融合方法. 而且采用本文的融合方法分类错误率下降得非常快. 另外，从表中也可以看出，组合不同的特征对识别率会产生一定的影响，其中组合 Gabor 特征与 Legendre 特征时，三种方法的识别率均好于其它特征组合下的结果. 因此，融合前的特征选择至关重要.

表 1 在不同的特征组合下各种方法的分类错误率比较

特征组合	分类错误率		
	串行融合方法	并行融合方法	本文的方法
X^G-X^L	0.1920(96)	0.1925(48)	0.1290(85)
X^G-X^P	0.2280(128)	0.2285(80)	0.2181(31)
X^G-X^Z	0.2295(116)	0.2290(74)	0.2255(20)
X^L-X^P	0.2400(102)	0.2410(61)	0.2160(34)
X^L-X^Z	0.2505(79)	0.2505(59)	0.2312(30)
X^P-X^Z	0.4760(51)	0.4785(31)	0.3295(30)

注: 表中括号中的数是在取得最优结果时的组合特征的维数.

3 高维小样本情况下的讨论

3.1 算法与理论

在模式识别领域，特别是人脸图像识别中，经常要处理高维小样本问题，这类问题中样本的总体散布矩阵(或协方差矩阵)一般来说是奇异的. 在这种情况下，如何求解样本特征空间 A 与 B(2.2 节)中矢量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 之间的典型相关特征，就是本节要讨论的主要问题.

CCA 作为一种重要的多元数据处理方法，虽然在诸多领域都有其应用，但应用也仅限于两个随机矢量所在的样本空间的协方差矩阵非奇异的情况[8~13]，文献[7]通过广义逆来解决奇异时的典型相关特征的求解问题，但其解并非理论意义上的精确解，而且在实际应用中有时也未必能取得好的效果. 正因为如此，也限制了它的应用. 我们的想法是，在不损失任何有效信息的前提下，将高维的原始样本的特征空间变换为低维的欧氏空间. 而在低维的欧氏空间内，总体的协方差矩阵是可逆的. 这样，求解典型投影矢量只需要在低维的欧氏空间内进行.

基于以上想法，我们将给出一般情况下的讨论. 假设上述样本空间 A 与 B 的协方差矩阵 $\boldsymbol{S}_{xx}$ 与 $\boldsymbol{S}_{yy}$ 均是奇异的. 由协方差矩阵的定义可知，此时 $\boldsymbol{S}_{xx}$ 与 $\boldsymbol{S}_{yy}$ 均是非负定的. 下面考虑典型投影矢量的求解问题.

设 $\xi_1, \xi_2, \cdots, \xi_p$ 与 $\eta_1, \eta_2, \cdots, \eta_q$ 分别表示矩阵 $\boldsymbol{S}_{xx}$ 与 $\boldsymbol{S}_{yy}$ 的标准正交的本征矢量，则

$$R^p = \mathrm{span}\{\xi_1, \xi_2, \cdots, \xi_p\};$$

$$R^q = \mathrm{span}\{\eta_1, \eta_2, \cdots, \eta_q\}.$$

令子空间

$$\Phi_x = \mathrm{span}\{\xi_1, \xi_2, \cdots, \xi_r\};$$

$$\Phi_y = \mathrm{span}\{\eta_1, \eta_2, \cdots, \eta_l\}.$$

其正交补空间分别为

$$\Phi_x^{\perp} = \mathrm{span}\{\xi_{r+1}, \xi_{r+2}, \cdots, \xi_p\};$$

$$\Phi_y^{\perp} = \mathrm{span}\{\eta_{l+1}, \eta_{l+2}, \cdots, \eta_q\}.$$

其中，$\boldsymbol{S}_{xx}$ 与 $\boldsymbol{S}_{yy}$ 的秩分别为 r 与 l，$\xi_1, \xi_2, \cdots, \xi_r$ 与 $\eta_1, \eta_2, \cdots, \eta_l$ 分别为 $\boldsymbol{S}_{xx}$ 与 $\boldsymbol{S}_{yy}$ 的非零本征值所对应的标准正交的本征矢量.

由上述约定可知，$R^p = \Phi_x \oplus \Phi_x^{\perp}$. 对 $\forall \alpha \in R^p$，α 可表示为 $\alpha = \alpha^* + \varphi$，其中 $\alpha^* \in \Phi_x$，$\varphi \in \Phi_x^{\perp}$. 因此，可定义映射 $f_1: \alpha \rightarrow \alpha^*$. 显然，$f_1$ 是从 R^p 到 Φ_x 的线性变换. 同理可定义映射 $f_2: \beta \rightarrow \beta^*$ ($R^q = \Phi_y \oplus \Phi_y^{\perp}$，$\beta = \beta^* + \psi$，$\beta^* \in \Phi_y$，$\psi \in \Phi_y^{\perp}$). 由线性代数的有关理论，不难得到如下引理.

引理 1. 设 $\boldsymbol{A}$ 为实对称矩阵，$\boldsymbol{x} \neq 0$，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = 0$ 当且仅当 $\boldsymbol{A}\boldsymbol{x} = 0$.

由此引理可以判定，$\Phi_x^{\perp}$ 为矩阵 $\boldsymbol{S}_{xx}$ 的零空间，$\Phi_y^{\perp}$ 为矩阵 $\boldsymbol{S}_{yy}$ 的零空间.

定理 4. 在上述线性变换 $f: \alpha \rightarrow \alpha^*$，$g: \beta \rightarrow \beta^*$ 下，有

$$J(\alpha^*, \beta^*) = J(\alpha, \beta).$$

证明. 由引理及 $\Phi_x^{\perp}$，$\Phi_y^{\perp}$ 的定义可知，

$$\alpha^{*\mathrm{T}}\boldsymbol{S}_{xx}\varphi = \varphi^{\mathrm{T}}\boldsymbol{S}_{xx}\varphi = \beta^{*\mathrm{T}}\boldsymbol{S}_{yy}\psi = \psi^{\mathrm{T}}\boldsymbol{S}_{yy}\psi = 0 \tag{10}$$

所以

$$\begin{aligned}\alpha^{\mathrm{T}}\boldsymbol{S}_{xx}\alpha &= \alpha^{*\mathrm{T}}\boldsymbol{S}_{xx}\alpha^* + 2\alpha^{*\mathrm{T}}\boldsymbol{S}_{xx}\varphi + \varphi^{\mathrm{T}}\boldsymbol{S}_{xx}\varphi \\ &= \alpha^{*\mathrm{T}}\boldsymbol{S}_{xx}\alpha^*\end{aligned} \tag{11}$$

同理，

$$\beta^{\mathrm{T}}\boldsymbol{S}_{yy}\beta = \beta^{*\mathrm{T}}\boldsymbol{S}_{yy}\beta^* \tag{12}$$

令

$$\boldsymbol{X} = (\boldsymbol{x}_1 - \mu_x, \boldsymbol{x}_2 - \mu_x, \cdots, \boldsymbol{x}_N - \mu_x),$$

$$\boldsymbol{Y} = (\boldsymbol{y}_1 - \mu_y, \boldsymbol{y}_2 - \mu_y, \cdots, \boldsymbol{y}_N - \mu_y),$$

其中 μ_x 与 μ_y 分别为样本空间 A 与 B 中样本均值，N 为样本总数. 则

$$\boldsymbol{S}_{xx} = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}, \quad \boldsymbol{S}_{yy} = \frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}.$$

由式(10)可推得

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$\varphi^{\mathrm T}\boldsymbol S_{xx}\varphi=\frac{1}{N}\varphi^{\mathrm T}\boldsymbol X\boldsymbol X^{\mathrm T}\varphi=\frac{1}{N}(\boldsymbol X^{\mathrm T}\varphi)^{\mathrm T}(\boldsymbol X^{\mathrm T}\varphi)=0$$

$$\Rightarrow\qquad \boldsymbol X^{\mathrm T}\varphi=0 \tag{13}$$

同理，有

$$\boldsymbol Y^{\mathrm T}\psi=0 \tag{14}$$

因此

$$\varphi^{\mathrm T}\boldsymbol S_{xy}\psi=\frac{1}{N}\varphi^{\mathrm T}\boldsymbol X\boldsymbol Y^{\mathrm T}\psi=\frac{1}{N}(\boldsymbol X^{\mathrm T}\varphi)^{\mathrm T}(\boldsymbol Y^{\mathrm T}\psi)=0 \tag{15}$$

利用式(13)和(14)，同理可得

$$\alpha^{*\mathrm T}\boldsymbol S_{xy}\psi=\varphi^{\mathrm T}\boldsymbol S_{xy}\beta^{*}=0 \tag{16}$$

因此，由式(15)和(16)得到

$$\alpha^{\mathrm T}\boldsymbol S_{xy}\beta=(\alpha^{*}+\varphi)^{\mathrm T}\boldsymbol S_{xy}(\beta^{*}+\psi)=\alpha^{*\mathrm T}\boldsymbol S_{xy}\beta^{*} \tag{17}$$

由式(11)，(12)，(17)，判别准则函数(4)变换为

$$\begin{aligned}J(\alpha,\beta)&=\frac{\alpha^{\mathrm T}\boldsymbol S_{xy}\beta}{(\alpha^{\mathrm T}\boldsymbol S_{xx}\alpha\cdot\beta^{\mathrm T}\boldsymbol S_{yy}\beta)^{1/2}}\\&=\frac{\alpha^{*\mathrm T}\boldsymbol S_{xy}\beta^{*}}{(\alpha^{*\mathrm T}\boldsymbol S_{xx}\alpha^{*}\cdot\beta^{*\mathrm T}\boldsymbol S_{yy}\beta^{*})^{1/2}}\\&=J(\alpha^{*},\beta^{*})\end{aligned}$$

证毕.

定理 4 说明了：典型投影矢量可分别在子空间 Φ_x 与 Φ_y 内选取，就判别准则函数(4)而言，不损失任何有效信息. 因此，模型 1 等价于

$$\text{模型 2}\begin{cases}\max J(\alpha,\beta)\\ \alpha^{\mathrm T}\boldsymbol S_{xx}\alpha=\beta^{\mathrm T}\boldsymbol S_{yy}\beta=1\\ \alpha_i^{\mathrm T}\boldsymbol S_{xx}\alpha=\beta_i^{\mathrm T}\boldsymbol S_{yy}\beta=0(i<k)\\ \alpha\in\Phi_x,\ \beta\in\Phi_y\end{cases}$$

下面讨论模型 2 的求解. 因为 $\dim\Phi_x=r$，$\dim\Phi_y=l$. 因此有 $\Phi_x\cong R^r$，$\Phi_y\cong R^l$，相应地同构映射可定义为 $g_1:\tilde\alpha\to\alpha$；$g_2:\tilde\beta\to\beta$. 即 $\alpha=\boldsymbol P\tilde\alpha$，$\beta=\boldsymbol Q\tilde\beta$. 其中

$$\boldsymbol P=(\xi_1,\xi_2,\cdots,\xi_r),\ \tilde\alpha\in R^r,\ \alpha\in\Phi_x;$$

$$\boldsymbol Q=(\eta_1,\eta_2,\cdots,\eta_l),\ \tilde\beta\in R^l,\ \beta\in\Phi_y.$$

在此同构映射下，准则函数 $J(\alpha,\beta)$ 变为

$$\begin{aligned}J(\alpha,\beta)&=\frac{\alpha^{\mathrm T}\boldsymbol S_{xy}\beta}{(\alpha^{\mathrm T}\boldsymbol S_{xx}\alpha\cdot\beta^{\mathrm T}\boldsymbol S_{yy}\beta)^{1/2}}\\&=\frac{\tilde\alpha^{\mathrm T}(\boldsymbol P^{\mathrm T}\boldsymbol S_{xy}\boldsymbol Q)\tilde\beta}{[\tilde\alpha^{\mathrm T}(\boldsymbol P^{\mathrm T}\boldsymbol S_{xx}\boldsymbol P)\tilde\alpha\cdot\tilde\beta^{\mathrm T}(\boldsymbol Q^{\mathrm T}\boldsymbol S_{yy}\boldsymbol Q)\tilde\beta]^{1/2}}\end{aligned}$$

定义函数：

$$\tilde J(\tilde\alpha,\tilde\beta)=\frac{\tilde\alpha^{\mathrm T}\tilde{\boldsymbol S}_{xy}\tilde\beta}{(\tilde\alpha^{\mathrm T}\tilde{\boldsymbol S}_{xx}\tilde\alpha\cdot\tilde\beta^{\mathrm T}\tilde{\boldsymbol S}_{yy}\tilde\beta)^{1/2}} \tag{18}$$

这里

$$\tilde{\boldsymbol S}_{xy}=\boldsymbol P^{\mathrm T}\boldsymbol S_{xy}\boldsymbol Q,\quad \tilde{\boldsymbol S}_{xx}=\boldsymbol P^{\mathrm T}\boldsymbol S_{xx}\boldsymbol P,\quad \tilde{\boldsymbol S}_{yy}=\boldsymbol Q^{\mathrm T}\boldsymbol S_{yy}\boldsymbol Q.$$

容易证明 $\tilde{\boldsymbol S}_{xx}$ 与 $\tilde{\boldsymbol S}_{yy}$ 均为正定矩阵且 $\tilde{\boldsymbol S}_{xy}^{\mathrm T}=(\boldsymbol P^{\mathrm T}\boldsymbol S_{xy}\boldsymbol Q)^{\mathrm T}=\boldsymbol Q^{\mathrm T}\boldsymbol S_{yx}\boldsymbol P=\tilde{\boldsymbol S}_{yx}$.

定理 5.　在上述同构映射 $\alpha=\boldsymbol P\tilde\alpha$，$\beta=\boldsymbol Q\tilde\beta$ 下，(α_0,β_0) 是准则函数 $J(\alpha,\beta)$ 的极值点当且仅当 $(\tilde\alpha_0,\tilde\beta_0)$ 是准则函数 $\tilde J(\tilde\alpha,\tilde\beta)$ 的极值点. 其中 $\tilde\alpha_0=\boldsymbol P^{\mathrm T}\alpha_0$，$\tilde\beta_0=\boldsymbol Q^{\mathrm T}\beta_0$.

定理 6.　(1) 设矩阵 $\boldsymbol P=(\xi_1,\xi_2,\cdots,\xi_r)$，$\alpha_i=\boldsymbol P\tilde\alpha_i$，$\alpha_j=\boldsymbol P\tilde\alpha_j$，则 α_i 与 α_j 关于 $\boldsymbol S_{xx}$ 共轭正交当且仅当 $\tilde\alpha_i$ 与 $\tilde\alpha_j$ 关于 $\tilde{\boldsymbol S}_{xx}$ 共轭正交；

(2) 设矩阵 $\boldsymbol Q=(\eta_1,\eta_2,\cdots,\eta_r)$，$\beta_i=\boldsymbol Q\tilde\beta_i$，$\beta_j=\boldsymbol Q\tilde\beta_j$，则 β_i 与 β_j 关于 $\boldsymbol S_{yy}$ 共轭正交当且仅当 $\tilde\beta_i$ 与 $\tilde\beta_j$ 关于 $\tilde{\boldsymbol S}_{yy}$ 共轭正交.

证明.　这里仅证明(1)，(2)可以类似地证明. 由于

$$\alpha_i^{\mathrm T}\boldsymbol S_{xx}\alpha_j=\tilde\alpha_i^{\mathrm T}(\boldsymbol P^{\mathrm T}\boldsymbol S_{xx}\boldsymbol P)\tilde\alpha_j=\tilde\alpha_i^{\mathrm T}\tilde{\boldsymbol S}_{xx}\tilde\alpha_j$$

所以

$$\alpha_i^{\mathrm T}\boldsymbol S_{xx}\alpha_j=\delta_{ij}\Leftrightarrow\tilde\alpha_i^{\mathrm T}\tilde{\boldsymbol S}_{xx}\tilde\alpha_j=\delta_{ij}.$$

故结论成立.　证毕.

在上述同构映射下，模型 2 又可以变换为

$$\text{模型 3}\begin{cases}\max\tilde J(\tilde\alpha,\tilde\beta)\\ \tilde\alpha^{\mathrm T}\tilde{\boldsymbol S}_{xx}\tilde\alpha=\tilde\beta^{\mathrm T}\tilde{\boldsymbol S}_{yy}\tilde\beta=1\\ \tilde\alpha_i^{\mathrm T}\tilde{\boldsymbol S}_{xx}\tilde\alpha=\tilde\beta_i^{\mathrm T}\tilde{\boldsymbol S}_{yy}\tilde\beta=0(i<k)\\ \tilde\alpha\in R^r,\ \tilde\beta\in R^l\end{cases}$$

根据定理 5 与定理 6，我们不难得出以下结论.

定理 7.　设

$$\tilde\alpha_1,\cdots,\tilde\alpha_d;\ \tilde\beta_1,\cdots,\tilde\beta_d\ (d\le\mathrm{rank}(\tilde{\boldsymbol S}_{xy}))$$

为模型 3 的最优解，则 $\alpha_1=\boldsymbol P\tilde\alpha_1,\cdots,\alpha_d=\boldsymbol P\tilde\alpha_d$；$\beta_1=\boldsymbol Q\tilde\beta_1,\cdots,\beta_d=\boldsymbol Q\tilde\beta_d$ 为典型投影矢量集.

至此，我们给出了样本的协方差矩阵 $\boldsymbol S_{xx}$ 与 $\boldsymbol S_{yy}$ 均奇异时，典型投影矢量集的求解. 即先求出 $\boldsymbol S_{xx}$ 与 $\boldsymbol S_{yy}$ 的非零本征值所对应的标准正交的本征矢量，令 $\boldsymbol P=(\xi_1,\xi_2,\cdots,\xi_r)$，$\boldsymbol Q=(\eta_1,\eta_2,\cdots,\eta_l)$；然后分别在 r 维与 l 维空间上计算样本的协方差矩阵 $\tilde{\boldsymbol S}_{xx}=\boldsymbol P^{\mathrm T}\boldsymbol S_{xx}\boldsymbol P$，$\tilde{\boldsymbol S}_{yy}=\boldsymbol Q^{\mathrm T}\boldsymbol S_{yy}\boldsymbol Q$ 以及它们之间的互协方差矩阵 $\tilde{\boldsymbol S}_{xy}=\boldsymbol P^{\mathrm T}\boldsymbol S_{xy}\boldsymbol Q$；最后按照 2.3 节中的算法步骤求解典型投影矢量集及典型相关特征矢量即可.

当 $\boldsymbol S_{xx}$ 与 $\boldsymbol S_{yy}$ 仅有一个奇异时，有类似的讨论，这里从略.

3.2　算法分析

由定理 7 所得的典型投影矢量集所构成的如下变换进行特征抽取

$$\boldsymbol Z=\begin{pmatrix}\boldsymbol W_x\\ \boldsymbol W_y\end{pmatrix}^{\mathrm T}\begin{pmatrix}\boldsymbol x\\ \boldsymbol y\end{pmatrix},$$

$$\begin{aligned}\boldsymbol W_x&=(\alpha_1,\alpha_2,\cdots,\alpha_d)=(\boldsymbol P\tilde\alpha_1,\boldsymbol P\tilde\alpha_2,\cdots,\boldsymbol P\tilde\alpha_d)\\&=\boldsymbol P(\tilde\alpha_1,\tilde\alpha_2,\cdots,\tilde\alpha_d)=\boldsymbol P\tilde{\boldsymbol W}_x;\end{aligned}$$

$$\boldsymbol W_y=(\beta_1,\beta_2,\cdots,\beta_d)=(\boldsymbol Q\tilde\beta_1,\boldsymbol Q\tilde\beta_2,\cdots,\boldsymbol Q\tilde\beta_d)$$

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　　http://www.cnki.net

$$= \boldsymbol{Q}(\beta_1, \beta_2, \cdots, \beta_d) = \boldsymbol{Q}\boldsymbol{W}_y.$$

则

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_x \\ \boldsymbol{W}_y \end{pmatrix} = \begin{pmatrix} \boldsymbol{P}\boldsymbol{W}_x \\ \boldsymbol{Q}\boldsymbol{W}_y \end{pmatrix} = \begin{pmatrix} \boldsymbol{P} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix} \begin{pmatrix} \boldsymbol{W}_x \\ \boldsymbol{W}_y \end{pmatrix}$$

上述变换分解为如下两个变换:

$$\begin{pmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\boldsymbol{y}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{P}^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \tag{19}$$

$$\boldsymbol{Z} = (\boldsymbol{W}_x^{\mathrm{T}} \quad \boldsymbol{W}_y^{\mathrm{T}}) \begin{pmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\boldsymbol{y}} \end{pmatrix} \tag{20}$$

在变换(19)中, $\tilde{\boldsymbol{x}} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}$, $\tilde{\boldsymbol{y}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{y}$, 特征空间 A 与 B 变成新的 r 维及 l 维的特征子空间. 由于变换矩阵 $\boldsymbol{P}$ 的列向量为矩阵 $\boldsymbol{S}_{xx}$ 的非零本征值所对应的本征矢量, 故该变换即为 K-L 变换, 且在 K-L 变换空间内, 样本的总体协方差矩阵为

$$\begin{aligned} & E\{(\tilde{\boldsymbol{x}} - \bar{\tilde{\boldsymbol{x}}})(\tilde{\boldsymbol{x}} - \bar{\tilde{\boldsymbol{x}}})^{\mathrm{T}}\} \\ & = E\{\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{P}\} \\ & = \boldsymbol{P}^{\mathrm{T}}E\{(\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathrm{T}}\}\boldsymbol{P} \\ & = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{S}_{xx}\boldsymbol{P} = \boldsymbol{S}_{xx}, \end{aligned}$$

同样, 变换矩阵 $\boldsymbol{Q}$ 也有类似的性质. 可见, 模型 3 确定的最优解 $\tilde{\alpha}_1, \tilde{\alpha}_2, \cdots, \tilde{\alpha}_d; \beta_1, \beta_2, \cdots, \beta_d$ 即为在两个 K-L 变换后的子空间内基于判别准则 $J(\tilde{\alpha}, \beta)$ 的典型投影矢量集.

通过以上分析, 我们可以归纳出奇异情况下典型相关特征矢量的抽取过程:

1. 利用 K-L 变换(以 $\boldsymbol{S}_{xx}$ 与 $\boldsymbol{S}_{yy}$ 为产生矩阵)分别将高维的原始样本压缩为 $\mathrm{rank}(\boldsymbol{S}_{xx})$ 与 $\mathrm{rank}(\boldsymbol{S}_{yy})$ 维;

2. 在变换后的特征空间内, 利用式(20)进行组合特征抽取.

以上我们讨论了当训练样本的协方差矩阵奇异时的典型相关特征的抽取问题, 即采取两阶段策略. 实际应用中, 在第一阶段利用 K-L 变换进行降维时, 有时根据问题的需要, 只要将原始的高维样本压缩为 k_1 维($k_1 < \mathrm{rank}(\boldsymbol{S}_{xx})$)与 k_2 维($k_2 < \mathrm{rank}(\boldsymbol{S}_{yy})$)即可. 也就是说在选取变换矩阵 $\boldsymbol{P}$ 与 $\boldsymbol{Q}$ 时, 可以将矩阵 $\boldsymbol{S}_{xx}$ 与 $\boldsymbol{S}_{yy}$ 较小的非零本征值对应的本征矢量舍去, 而只保留其对应本征值较大的本征矢量. 因过小的本征值有时会含有较多的干扰信息, 不利于分类.

3.3 实验 2

FERET 人脸图像数据库由美国国防部发起建立, 其初衷是想开发一个自动人脸识别系统, 以应用于各种安全检测. 现已成为国际上通用的对人脸识别算法进行评估的一个标准数据库[15]. 其最新的人脸数据库已包括 14051 幅, 分辨率为 384×256 的灰度图像. 该库对人脸的姿态和表情有着严格的规定.

本实验是在 FERET 人脸库的一个子库上进行. 该子库包含 146 个人的 584 幅图像(每人 4 幅. 其中的两幅以 fa 标识, 另两幅以 fb 标识), 是从 FERET 人脸库中挑出的一些具有人眼坐标信息的正面人脸图像. 图 1 给出了部分典型图像.

图 1 FERET 人脸图像数据库中的部分典型图像

图 2 FERET 中的原始图像经过预处理后的典型图像

该库中的图像包含了人的上半身, 所以需要进行预处理, 以获取仅包含人脸部分的图像. 预处理采用基于 CSU(Colorado State University)的人脸验证评估系统中提供的主体算法[16]. 主要的预处理工作包括几何校正(geometric normalization)、图像掩模(masking)、直方图均衡化(histogram equalization)、像素灰度值归一化(pixel normalization)等. 经过预处理后将原图像变成分辨率为 150×130 的灰度图像, 其部分典型人脸图像如图 2 所示.

本实验采用每人的两幅 fa 图像作为训练样本, 余下的两幅 fb 图像作为测试样本, 这样训练样本和测试样本的总数均为 292.

通过分辨率为 150×130 的原始图像组成第一类训练样本空间 $A = \{\boldsymbol{x} \mid \boldsymbol{x} \in R^{19500}\}$; 利用正交小波对所有原始图像进行 3 次小波变换, 获得分辨率为 19×17 的低频图像, 构成第二类训练样本空间 $B = \{\boldsymbol{y} \mid \boldsymbol{y} \in R^{323}\}$. 将这两组特征进行组合. 因为 $\boldsymbol{S}_{xx}$ 与 $\boldsymbol{S}_{yy}$ 均奇异, 所以我们首先利用第 3 节的讨论, 通过 K-L 变换将训练样本空间 A 中的 19500 维的特征矢量压缩为 230 维的特征矢量, 组成一个新的特征

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

空间 $\boldsymbol{A}=\{\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{x}}\in R^{230}\}$；将训练样本空间 B 中的 323 维的特征矢量压缩为 60 维的特征矢量，组成另一个特征空间 $\boldsymbol{B}=\{\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{y}}\in R^{60}\}$. 然后，在 K-L 变换后的两个特征子空间中，使用本文提出的方法抽取它们组合的典型相关特征，分别采用最小距离分类器和最近邻分类器进行分类，分类结果如表 2 所示.

另外，为了说明本文提出的特征融合方法的有效性，在表 3 中，我们也给出了基于单一特征的 Eigenfaces 方法与 Fisherfaces 方法在原始人脸图像上的分类结果[17, 18].

表 2　本文方法在最小距离分类器和最近邻分类器下的分类错误数

投影轴数	最小距离下的分类错误数	最近邻下的分类错误数
20	42	36
22	35	32
24	28	29
26	30	24
28	30	22
30	26	19
32	21	16
34	18	15
36	18	12
38	20	13
40	18	15
42	15	13
44	16	12
46	13	10
48	11	9
50	10	7
52	12	8
54	10	9
56	13	8
58	15	9
60	14	7

表 3　**Eigenfaces 法、Fisherfaces 法**以及本文方法在两种分类器下分类结果的比较

方 法	最小距离下的识别率	最近邻下的识别率
Eigenfaces	0.9349(147)	0.9486(76)
Fisherfaces	0.9418(146)	0.9349(67)
本文方法	0.9658(49)	0.9760(50)

注：表中括号中的数是在取得最优结果时的鉴别特征的维数.

从表 2 和表 3 可以看出，本文算法在最小距离分类器和最近邻分类器下均取得了比较好的分类结果. 当维数在 45～60 变动时，识别率均保持在 95％以上，其最优识别率分别达到了 96.58％与 97.6％，优于基于单一特征的 Eigenfaces 方法和 Fisherfaces 方法. 实验结果再次表明了组合后的典型相关特征具有较强的鉴别能力，是一种有效的特征融合方法.

4　算法的有效性分析

在模式识别理论中，特征抽取的一般原则是所抽取的特征之间的统计相关性越小越好，最好是抽取不相关的特征[19]. 由定理 1 和定理 3 知道，利用典型相关分析所抽取的典型相关特征之间是不相关的，利用变换(3)所抽取的组合鉴别特征之间也是不相关的. 因此所获得的投影变换是最优的.

另外，实验 1 中的结果表明，与文献[4, 5]两种特征融合方法相比，本文提出的特征融合方法在特征抽取的速度上占有一定的优势. 通常，利用代数方法进行特征抽取所花费的时间，很大程度上取决于投影矢量集的求解过程，即本征值与本征矢量的求解. 当矩阵的阶数很大时，本征值与本征矢量的计算是非常耗时的. 若同一模式的两组特征矢量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 分别是 p 维和 q 维的，文献[4]的投影矢量要在 $p+q$ 维的实向量空间上进行，文献[5]要在 $\max(p, q)$ 维的复向量空间进行，而本文的方法则只需在 $\min(p, q)$ 维的实向量空间上进行. 当 p 与 q 差别较大，这种计算上的优势(速度)是非常明显的. 例如在实验 1 中，当组合特征为 X^{G} 与 X^{Z} 时，文献[4, 5]及本文算法分别需要求解一个 286, 256 及 30 阶矩阵的本征值和本征矢量.

此外，在特征抽取和分类识别中，基于 Fisher 准则的线性鉴别分析是公认的最有效的方法之一. 可以从理论上证明，Fisher 线性鉴别分析是典型相关分析的一种特殊情况，通过 Fisher 鉴别分析求解的分类问题可以转化为 CCA 的算法予以解决(我们将另文讨论). 从这个意义上来说，典型相关分析用于模式识别更具有一般性，也更具发展潜力.

5　结　论

本文第一次将 CCA 的思想用于特征级融合及图像识别领域，提出了一种新的组合特征抽取的方法. 该方法巧妙地将两组特征之间的相关性特征作为判别信息，既达到了信息融合之目的，又有效地消除了冗余信息，为两组特征融合用于分类提供了新的思路. 本文不仅给出了 CCA 用于图像识别的理论和方法，更重要的是从理论上推广了 CCA 的应用范围，解决了当两组特征矢量构成的总体协方差矩阵奇异时，典型相关特征的求解问题，使之适合于高维小样本的情形. 在两个数据库上的实验结果表明了，所抽取的典型相关特征不仅可以实现原始特征维数的压缩，而且具有良好的分类性能.

本文提出的方法为特征级信息融合提供了一条

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

行之有效的途径.进一步的工作是,从理论上进一步完善该方法,探讨 CCA 与 Fisher 线性鉴别分析的内在联系,拓宽其研究领域.

最后,值得一提的是,本文第 3 节得到的结论是对 CCA 理论的完善和发展,完全适用于利用 CCA 解决问题的其它领域.

参 考 文 献

1 Huang Y.S., Suen C.Y.. A method of combining multiple experts for the recognition of unconstrained handwritten numerals. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 7(1): 90~94

2 Constantinidis A.S., Fairhurst M.C., Rahman A.F.R.. A new multi expert decision combination algorithm and its application to the detection of circumscribed masses in digital mammograms. Pattern Recognition, 2001, 34(8): 1528~1537

3 Jing X.Y., Zhang D., Yang Z.Y.. Face recognition based on a group decision making combination approach. Pattern Recognition, 2003, 36(7): 1675~1678

4 Liu C.J., Wechsler H.. A shape and texture based enhanced Fisher classifier for face recognition. IEEE Transactions on Image Processing, 2001, 10(4): 598~608

5 Yang J., Yang J.Y.. Generalized K-L transform based combined feature extraction. Pattern Recognition, 2002, 35(1): 295~297

6 Yang J., Yang J.Y., Zhang D., Lu J.F.. Feature fusion: Parallel strategy vs. serial strategy. Pattern Recognition, 2003, 36(6): 1369~1381

7 Zhang X.T., Fang K.T.. Multivariate Statistical Introduction. Beijing: Science Press, 1999(in Chinese)
(张尧庭, 方开泰. 多元统计分析引论. 北京: 科学出版社, 1999)

8 Sun W.S., Chen L.X.. Multivariate statistical analysis. Beijing: Higher Education Press, 1994(in Chinese)
(孙文爽, 陈兰祥. 多元统计分析. 北京: 高等教育出版社, 1994)

9 Borga M.. Learning multidimensional signal processing. Department of Electrical Engineering[Ph.D. dissertation]. Linköping University, Linköping, Sweden, 1998

10 Yu S.J.. Direct blind channel equalization via the programmable canonical correlation analysis. Signal Processing, 2001, 81(8): 1715~1724

11 Choi H.C., King R.W.. Speaker adaptation through spectral transformation for HMM based speech recognition. In: Proceedings of the IEEE International Symposium on Speech, Image Processing and Neural Networks, Hong Kong, 1994, 2: 686~689

12 Melzer T., Reiter M., Bischof H.. Appearance models based on kernel canonical correlation Analysis. Pattern Recognition, 2003, 36(9): 1961~1971

13 Hotelling H.. Relations between two sets of variates. Biometrika, 1936, 28: 321~377

14 Hu Z.S., Lou Z., Yang J.Y., Liu K., Sun J.Y.. Handwritten digit recognition basis on multi classifier combination. Chinese Journal of Computers, 1999, 22(4): 369~374(in Chinese)
(胡钟山, 娄 震, 杨靖宇, 刘 克, 孙靖夷. 基于多分类器组合的手写体数字识别. 计算机学报, 1999, 22(4): 369~374)

15 Phillips P.J., Moon H.J., Rizvi S.A., Rauss P.J.. The FERET evaluation methodology for face recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(10): 1090~1104

16 Bolme D.S., Beveridge J.R., Teixeira M., Draper B.A.. The CSU face identification evaluation system: Its purpose, features, and structure. In: Proceedings of the 3rd International Conference on Computer Vision Systems(ICVS), Graz, Austria, 2003, 304~313

17 Turk M., Pentland A.. Face recognition using Eigenfaces. In: Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 1991, 586~591

18 Belhumeur P.N. *et al*.. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711~720

19 Jin Z., Yang J.Y., Tang Z.M., Hu Z.S.. A theorem on the uncorrelated optimal discriminant vectors. Pattern Recognition, 2001, 34(7): 2041~2047

SUN Quan Sen, born in 1963, Ph.D. candidate, associate professor. His current research interests include pattern recognition, image processing, computer vision and data fusion.

ZENG Sheng Gen, bron in 1977, Ph.D.. His current interests include pattern recognition, image processing and remote sensing image.

HENG Pheng Ann, bron in 1961, Ph.D., professor. His research interests include virtual reality applications in medicine, scientific visualization, 3D medical imaging.

XIA De Shen, bron in 1941, Ph.D., professor and Ph.D. supervisor. His research interests include image processing, remote sensing, medical image analysis and pattern recognition.

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Background

This work was supported by the Research Grants Council of the Hong Kong Special Administrative Region under Earmarked Research Grant (project no. CUHK4185/00E).

Information fusion technology is one of the emerging technologies of data processing. Among the three levels of information fusion (pixel level, feature level, and decision level), the advantage of the feature level fusion is obvious. Different feature vectors extracted from the same pattern always reflects the different characteristic of patterns. By optimizing and combining these different features, it not only keeps the effective discriminant information of multi feature, but also eliminates the redundant information to certain degree. This is especially important to classification and recognition.

Canonical Correlation Analysis (CCA) is one of the statistical methods dealing with the mutual relationships between two random vectors, and it has the same importance as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in Multivariate Statistical Analysis. It is one of the valuable multi data processing methods. In recent years, CCA has been applied to several fields such as signal processing, computer vision and sound recognition, and it got delightful development. So far there is no the relative research of CCA used in image recognition.

In this paper, a new strategy of feature level fusion is proposed adopting the idea of CCA and used in combination feature extraction. Experimental results on two database show that the proposed algorithm is efficient and robust. At the same time, the experimental results also show that this algorithm cannot only realize the decrease of primitive feature dimensions, but also be of good classification performance, reflecting the essential feature of the images. The method in this paper is a good approach for information fusion on feature level, and it shows the significances on theory and application.

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net